# Comparison of family 12 glycoside hydrolases and recruited substitutions important for thermal stability

MATS SANDGREN,[1] PETER J. GUALFETTI,[2] ANDREW SHAW,[2] LAURIE S. GROSS,[2] MAE SALDAJENO,[2] ANTHONY G. DAY,[2] T. ALWYN JONES,[1] AND COLIN MITCHINSON[2]

[1]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, SE-75124 Uppsala, Sweden
[2]Genencor International Inc., Palo Alto, California 94304, USA

(RECEIVED October 18, 2002; FINAL REVISION January 13, 2003; ACCEPTED January 13, 2003)

**Abstract**

As part of a program to discover improved glycoside hydrolase family 12 (GH 12) endoglucanases, we have studied the biochemical diversity of several GH 12 homologs. The *H. schweinitzii* Cel12A enzyme differs from the *T. reesei* Cel12A enzyme by only 14 amino acids (93% sequence identity), but is much less thermally stable. The bacterial Cel12A enzyme from *S. sp. 11AG8* shares only 28% sequence identity to the *T. reesei* enzyme, and is much more thermally stable. Each of the 14 sequence differences from *H. schweinitzii* Cel12A were introduced in *T. reesei* Cel12A to determine the effect of these amino acid substitutions on enzyme stability. Several of the *T. reesei* Cel12A variants were found to have increased stability, and the differences in apparent midpoint of thermal denaturation ($T_m$) ranged from a 2.5°C increase to a 4.0°C decrease. The least stable recruitment from *H. schweinitzii* Cel12A was A35S. Consequently, the A35V substitution was recruited from the more stable *S. sp. 11AG8* Cel12A and this *T. reesei* Cel12A variant was found to have a $T_m$ 7.7°C higher than wild type. Thus, the buried residue at position 35 was shown to be of critical importance for thermal stability in this structural family. There was a ninefold range in the specific activities of the Cel12 homologs on o-NPC. The most and least stable *T. reesei* Cel12A variants, A35V and A35S, respectively, were fully active. Because of their thermal tolerance, *S. sp. 11AG8* Cel12A and *T. reesei* Cel12A variant A35V showed a continual increase in activity over the temperature range of 25°C to 60°C, whereas the less stable enzymes *T. reesei* Cel12A wild type and the destabilized A35S variant, and *H. schweinitzii* Cel12A showed a decrease in activity at the highest temperatures. The crystal structures of the *H. schweinitzii*, *S. sp. 11AG8*, and *T. reesei* A35V Cel12A enzymes have been determined and compared with the wild-type *T. reesei* Cel12A enzyme. All of the structures have similar Cα traces, but provide detailed insight into the nature of the stability differences. These results are an example of the power of homolog recruitment as a method for identifying residues important for stability.

**Keywords:** Thermal stability; cellulase; endoglucanase; homolog; protein crystal structure

Reprint requests to: Alwyn Jones, Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, P.O. Box 596, S-751 24 Uppsala, Sweden; e-mail: alwyn@xray.bmc.uu.se; fax: 46-18-536971; or Colin Mitchinson, Genecor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304, USA; e-mail: cmitchinson@genencor.com; fax: (650) 845-6510.

*Abbreviations:* CD, circular dichroism; cd, catalytic domain; GH, glycoside hydrolase; HIC, hydrophobic interaction chromatography; *F. javanicum*, *Fusarium javanicum*; *G. roseum*, *Gliocladium roseum*; *H. schweinitzii*, *Hypocrea schweinitzii*; mme, mono-methyl-ether; MR, molecular replacement; NAG, N-acetyl-glucosamine; NCS, noncrystallographic symmetry; oNPC, o-Nitrophenyl β-cellobioside; PEG, polyethylene glycol; RMSD, root-mean-square deviation; *S. sp 11AG8*, *Streptomyces sp. 11AG8*; $T_m$, the mid-point of thermal denaturation; *T. reesei*, *Trichoderma reesei*; *T. koningii*, *Trichoderma koningii*.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.0237703.

Bacterial and fungal cellulases have been intensively studied for many years. They are used widely in the detergent, textile, and food industries. Cellulose and hemicellulose account for more than half of the organic carbon in the biosphere, and there is continuing interest in the potential use of cellulases in the conversion of cellulosic biomass to fermentable sugars. Cellulases are glycoside hydrolases found in at least 12 families of this very large group of enzymes (Henrissat and Davies 2000), and consist of cellobiohydrolases (or exoglucanases, EC 3.2.1.91) and endoglucanases (EC 3.2.1.4). Glycoside hydrolase family 12 (GH 12) members hydrolyse the β-1,4-glycosidic bond in cellulose via a double displacement reaction and a glycosyl-enzyme inter-

 

mediate that results in retention of the anomeric configuration in the product (Schulein 1997; Birsan et al. 1998). Structural studies on GH 12 members reveal a compact β-sandwich structure that is curved to create an extensive cellulose-binding site on the concave face of the β-sheet (Sulzenbacher et al. 1999; Fig. 1A,B). Structures from both bacterial (Sulzenbacher et al. 1997; Crennell et al. 2002) and fungal (Sandgren et al. 2001; Khademi et al. 2002) sources have now been determined, and these provide the structural framework for the whole family. As part of a program to discover cellulases with improved properties, many novel GH 12 endoglucanases have been cloned, sequenced, and expressed (van Solingen et al. 2001; Goedegebuur et al. 2002). The Cel12A enzyme from the alkalophilic bacterium *S. sp. 11AG8* was found to be more stable to thermal inactivation than the homologous enzyme from the filamentous fungus *T. reesei*, and was identified as a good candidate for industrial applications (van Solingen et al. 2001). The present work is part of a study of the biochemical and structural diversity of this sequence-diverse family to understand the sequence-structure-function relationships within the GH 12 proteins.

In this study, we have measured the stability and activity of several GH 12 homologs and point mutants of *T. reesei* Cel12A, in an attempt to identify residues important for thermal stability. Significant biochemical diversity was seen among the GH 12 homologs (Table 1a). Notably, the enzyme from the fungus *H. schweinitzii* (Goedegebuur et al. 2002) differs from the *T. reesei* Cel12A enzyme at only 14 residues, but is significantly less thermally stable. We have systematically introduced these 14 differences into the *T. reesei* Cel12A enzyme by site-directed mutagenesis, and examined their effects on the thermal stability of the enzyme. Several mutations recruited from *H. schweinitzii* affected stability, and A35S (all numbering is based on the *T. reesei* Cel12A sequence unless otherwise noted) was found to be the most destabilizing. Consequently, the A35V mutation was recruited from the more stable *S. sp. 11AG8* Cel12A enzyme and was found to be stabilizing. Thus, this process has identified buried residue 35 as being of critical importance. The large differences in stability between the enzymes prompted us to determine the crystal structure of two related GH 12 enzymes, the fungal Cel12A from *H. schweinitzii* and cd of the bacterial Cel12A from the *S. sp. 11AG8*, and of the most stable *T. reesei* Cel12A mutant, A35V. Comparison of these three structures was used to provide detailed rationale for the observed stability differences.

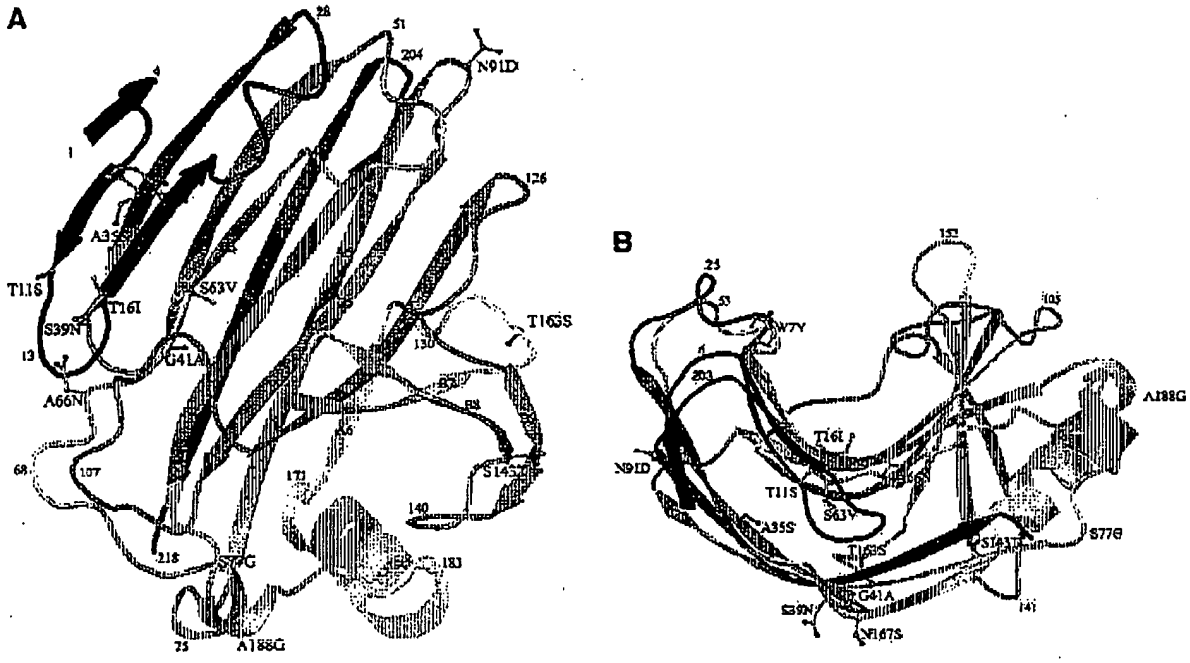


Figure 1. Schematic ribbon diagram. Top, (*A*) and side (*B*) views of the *H. schweinitzii* Cel12A crystal structure, color ramped according to residue number, starting with red at the amino terminus and ending with blue at the carboxyl terminus of the structure. The two β-sheets in the structure are labeled A and B, with the individual strands labeled (A1–A6 and B1–B9) according to their positions in the two β-sheets. The structures have side chains drawn for the 14 residues that differ from the *T. reesei* Cel12A protein sequence. Figures 1 and 3 were prepared using O (Jones et al. 1991), and rendered with Molray (Harris and Jones 2001).

**Table 1a.** *Thermal denaturation data and relative specific enzyme activity*

| GH12 homolog | $\Delta T_m$[a] | $T_m$ (°C)[b] | Fit error[c] | Activity[d] |
|---|---|---|---|---|
| *T. reesei* Cel12A | 0.0 | 54.4 | 0.19 | 1.00 |
| *H. schweinitzii* Cel12A | −5.2 | 49.2 | 0.08 | 1.08 |
| *S. sp. 11AG8* Cel12A cd | 11.3 | 65.7 | 0.06 | 3.84 |
| *S. sp. 11AG8* Cel12A fl | 12.4 | 66.8 | 0.03 | 3.78 |
| *T. koningii* Cel12A | 4.1 | 58.5 | 0.76 | 1.00 |
| *G. roseum* Cel12C | −8.5 | 45.9 | 0.17 | 0.42 |
| *F. javanicum* Cel12A | −0.3 | 54.1 | 0.07 | ND |

[a] $\Delta T_m$ values are relative to *T. reesei* Cel12A wild type.
[b] The thermal denaturation experiments were performed at 217 nm, in 0.05 M Bis-Tris propane, 0.05 M ammonium acetate (pH 8.0) by increasing the temperature from 30 to 90°C with data collected every 2 degrees. The midpoint of the transition ($T_m$) is an apparent value because the thermal denaturation is not reversible.
[c] $T_m$ fit error corresponds to one standard deviation.
[d] The specific enzyme activity was measured in 10 min incubation with oNPC at 40°C (pH 5.5). It is expressed as a mole specific activity relative to *T. reesei* Cel12A wild type. The standard deviations are 10% or less.

## Results

### *Stability determination*

The CD signal at 217 nm as a function of temperature was collected for all of the homologs and mutants and fitted to two-state models (Fig. 2A). The parameters from those fits were used to calculate a curve representing the apparent fraction of unfolded protein, $F_{app}$, and the fitted curve for each sample (Fig. 2B). The apparent $T_m$ for each sample is listed in Table 1a. As has been reported previously (Ward et al. 1993), the recombinant *T. reesei* Cel12A proteins expressed in *Aspergillus niger* were hyperglycosylated, as was *H. schweinitzii* Cel12A (data not shown). All stability determinations reported were on proteins deglycosylated with endoglucanase H. For the wild-type *T. reesei* Cel12A, this resulted in a form with the same $T_m$ as the native protein expressed in *T. reesei* (data not shown). This is consistent with the direct observation of an identical single NAG residue on Asn 164 in the native *T. reesei* Cel12A (Sandgren et al. 2001) and in the deglycosylated recombinant A35V variant and *H. schweinitzii* Cel12A.

Compared with the *T. reesei* Cel12A enzyme, which has a $T_m$ of 54.4°C, the *H. schweinitzii* enzyme has a $T_m$ that is 5.2°C lower, whereas the *S. sp. 11AG8* Cel12A homolog has a $T_m$ that is 12.6°C higher. The stability ranking by $T_m$ (Table 1a) was the same as that seen by thermal inactivation rates at pH8 (data not shown). The variants recruited from the *H. schweinitzii* Cel12A enzyme have $T_m$ changes ranging from −4.0°C to +2.5°C for the most stable mutant. The $\Delta T_m$ values given in Table 1b have a standard deviation of 0.2°C or less, and $T_m$ differences of <0.5°C were not considered significant. The most significant changes in stability occur within the first 63 amino acids. The A35S mutant has the largest decrease in stability ($\Delta T_m$ = −4.0°C), whereas mutating the same residue to a valine produces the largest increase in stability ($\Delta T_m$ = +7.7°C).

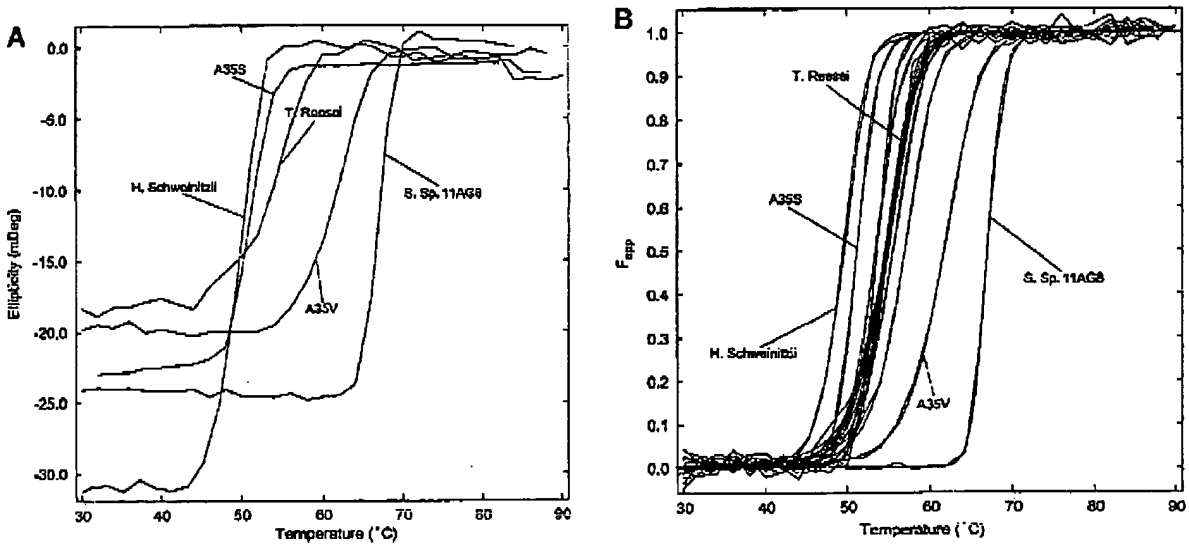


Figure 2. The raw thermal denaturation data (*A*) is shown for a subset of proteins examined. These data and similar traces for all of the proteins listed in Tables 1a and 1b were fit to two-state models and used to produce a fraction-apparent unfolded, $F_{app}$, plot (*B*). By this convention, 0 represents native and 1 represents unfolded state for the protein. Only a subset of the data sets are labeled. The protein concentration for the experiments was between 10 and 20 μM.

**Table 1b.** *Thermal denaturation data for* T. reesei *Cel12A variants*

| *T. reesei* Cel12A variant | $\Delta T_m$[a] | $T_m$ (°C)[b] | Fit error[c] |
|---|---|---|---|
| A35V | 7.7 | 62.1 | 0.07 |
| W7Y | −1.0 | 53.4 | 0.24 |
| T11S/T16I | 1.1 | 55.5 | 0.13 |
| A35S | −4.0 | 50.4 | 0.14 |
| S39N | 0.5 | 54.9 | 0.17 |
| G41A | 2.5 | 56.9 | 0.11 |
| S63V | −0.8 | 53.6 | 0.11 |
| A66N | 0.1 | 54.5 | 0.10 |
| S77G | 0.1 | 54.5 | 0.09 |
| N91D | 0.5 | 54.9 | 0.17 |
| S143T | 0.5 | 54.9 | 0.12 |
| T163S | 0.3 | 54.7 | 0.07 |
| N167S | 0.2 | 54.6 | 0.10 |
| A188G | 0.5 | 54.9 | 0.17 |

[a] $\Delta T_m$ values are relative to *T. reesei* Cel12A wild type.
[b] The thermal denaturation experiments were performed at 217 nm, in 0.05 M Bis-Tris propane, 0.05 M ammonium acetate (pH 8.0), by increasing the temperature from 30 to 90°C with data collected every 2 degrees. The midpoint of the transition ($T_m$) is an apparent value because the thermal denaturation is not reversible.
[c] $T_m$ Fit error corresponds to one standard deviation.

### *Relative enzyme activity*

The expression of all of the Cel12 homologs and variants in *A. niger* was detected by activity on derivatized cellulose (Goedegebuur et al. 2002), and o-Nitrophenyl β-cellobioside (oNPC) assays were used to monitor the purifications. Thus, all of these proteins are active enzymes. There was a ninefold difference in the specific activities of the Cel12 homologs on oNPC (Table 1a). The *T. reesei* Cel12A variants with the greatest and least stability (A35V and A35S, respectively) both showed ~30% higher activity than wild type (Fig. 3).

The activity of selected Cel12 enzymes as a function of temperature (Fig. 3) was determined, and corresponded to apparent activation energies between 10 and 11.5 kcal/mole in all cases. The Cel12 enzymes with greatest thermal stability, Cel12A *S. sp. 11AG8* and *T. reesei* Cel12A variant A35V, showed a continual increase in activity over the full temperature range from 25°C to 60°C (Fig. 3). The less stable homologs, *T. reesei* Cel12A and *H. schweinitzii* Cel12A, and the destabilized *T. reesei* Cel12A variant A35S, showed a decrease in activity at the highest temperature, presumably due to thermal inactivation.

### *Structure determination*

To better understand the structural basis for thermal stability in this family of enzymes, the structures of *H. schweinitzii*

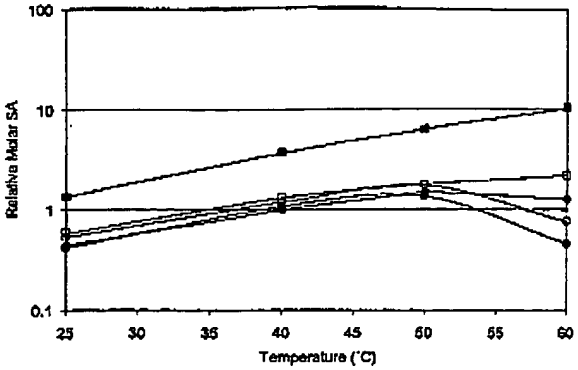


**Figure 3.** The temperature dependence of enzyme activity was assayed by hydrolysis of the chromogenic substrate oNPC in 0.05 M Bis-Tris propane, 0.05 M ammonium acetate (pH 5.5) over 10 min at 25°C, 40°C, 50°C, and 60°C. Data are shown for *S. sp. 11AG8* Cel12A cd (■), *T. reesei* Cel12A variant A35V (□), *T. reesei* Cel12A (◆), *T. reesei* Cel12A variant A35S (○), and *H. schweinitzii* Cel12A (●). It is expressed as a molar specific activity relative to *T. reesei* Cel12A at 40°C.

Cel12A, the *T. reesei* Cel12A A35V mutant, and *S. sp. 11AG8* Cel12A cd have been determined and refined to reasonably high resolution (Tables 2a, 2b). The protein structures have the expected fold of a GH 12 enzyme. They consist mainly of 15 β-strands building up 2 β-sheets, A and

**Table 2a.** *X-ray data collection and processing statistics*

| GH 12 dataset | *T. reesei* A35V | *S. sp. 11AG8* wild type | *H. schweinitzii* wild type |
|---|---|---|---|
| Collected | ESRF[a] ID14 EH1 | Rotating Anode | ESRF ID14 EH4 |
| Detector | MAR CCD 165 | Raxis II | ADSC Quanta 4r |
| Wavelength Å | 0.93 | 1.54 | 0.98 |
| Oscillation range (°) | 0.5 | 0.5 | 0.5 |
| Space group | $P2_1$ | $P2_12_12_1$ | $P2_1$ |
| Cell parameters (Å) | | | |
| a = | 68.3 | 65.2 | 62.5 |
| b = | 71.3 | 54.6 | 77.5 |
| c = | 119.3 | 62.6 | 83.4 |
| β = | 91.5° | | 98.5° |
| Resolution range (Å) | 25–1.5 | 29–1.5 | 30–1.7 |
| Resolution range outer shell | 1.53–1.50 | 1.53–1.50 | 1.73–1.70 |
| No. of observed reflections | 490,560 | 88,448 | 346,287 |
| No. of unique reflections | 172,895 | 26,559 | 85,680 |
| Average multiplicity | 2.8 | 3.3 | 4.0 |
| Completeness (%)[b] | 94.4 (94.3) | 72.9 (63.5) | 98.9 (84.6) |
| $R_{merge}$ (%)[c] | 3.7 (38.1) | 5.4 (16.1) | 9.7 (12.6) |
| $\langle I/\sigma(I)\rangle$ | 24.9 (2.3) | 15.4 (8.6) | 14.0 (5.7) |

[a] European Synchrotron Radiation Facility (ESRF), Grenoble, France.
[b] Numbers in parentheses are for their highest resolution bins.
[c] $R_{merge} = \Sigma_{hkl}\Sigma_i|I - \langle I\rangle|/\Sigma_{hkl}\Sigma_i|I|$.

B, A consisting of 6 β-strands (A1–A6) and B of 9 β-strands (B1–B9) stacked on top of one another, forming a β-sandwich (Fig. 1A,B). The concave surface of the larger β-sheet B produces a large substrate-binding cleft across one face of the enzyme (Sulzenbacher et al. 1997). From structural comparisons with GH 12 homologs with known structure in which the catalytic residues have been identified (Zechel et al. 1998; Sulzenbacher et al. 1999), the identities of the catalytic nucleophile and the acid/base have been deduced. These correspond to glutamyl residues 116 and 200, respectively, in *T. reesei* Cel12A, an identification supported by site-directed mutagenesis (Okada et al. 2000).

### H. schweinitzii *Cel12A structure*

The fungal Cel12A enzyme from *H. schweinitzii* crystallizes with four NCS-related molecules in the asymmetric unit. It differs at only 14 residues in the protein sequence (Fig. 1; Table 1b) from the fungal *T. reesei* Cel12A enzyme, and there are no insertions or deletions. The complete set of Cα atoms from the four NCS molecules in the *H. schweinitzii* Cel12A structure and the six NCS molecules in *T. reesei* Cel12A structure can be superimposed with pair-wise RMSD in the range from 0.3 to 0.5 Å. Some of the biggest main-chain differences between the two structures can be found in the two loops corresponding to residues 11–16 and 37–43, connecting β-strands B1 to B2 and A2 to A3 in the structure (Fig. 1A,B).

The *H. schweinitzii* Cel12A enzyme has a 5.2°C lower $T_m$ than the *T. reesei* enzyme (Table 1a). Most of the 14 differences in sequence are located on the protein surface and distributed over the whole molecule (Fig. 1A,B). The side chains at many of these sites point out into the surrounding solution. All of these substitutions (except N91D) are from a neutral to another neutral amino acid, and most have little or no effect on the $T_m$ of the enzyme (with $\Delta T_m$'s in the range from 0.1°C to 0.5°C) (Table 1b). There is a clustering of substitutions on the first β-strands in the structure, A2–A3 and B1–B3 (Fig. 1A,B), in which some of the individual substitutions have a large effect on the $T_m$ of the *T. reesei* Cel12A enzyme. This region also has some of the largest structural differences between the structures of the two fungal enzymes. The substitution in this region that has the biggest effect on the $T_m$ is the alanine to serine substitution at residue 35 on β-strand A2 (Fig. 4A). This substitution causes a reduction of the $T_m$ by 4.0°C when introduced into *T. reesei* Cel12A (Table 1b). The likely explanation for this is the introduction of a hydrophilic residue in the hydrophobic environment at the edge of the two β-sheets, disrupting the hydrophobic interactions with the side chains of the three surrounding hydrophobic residues (F10, V17, and L38, with interatomic separations in the

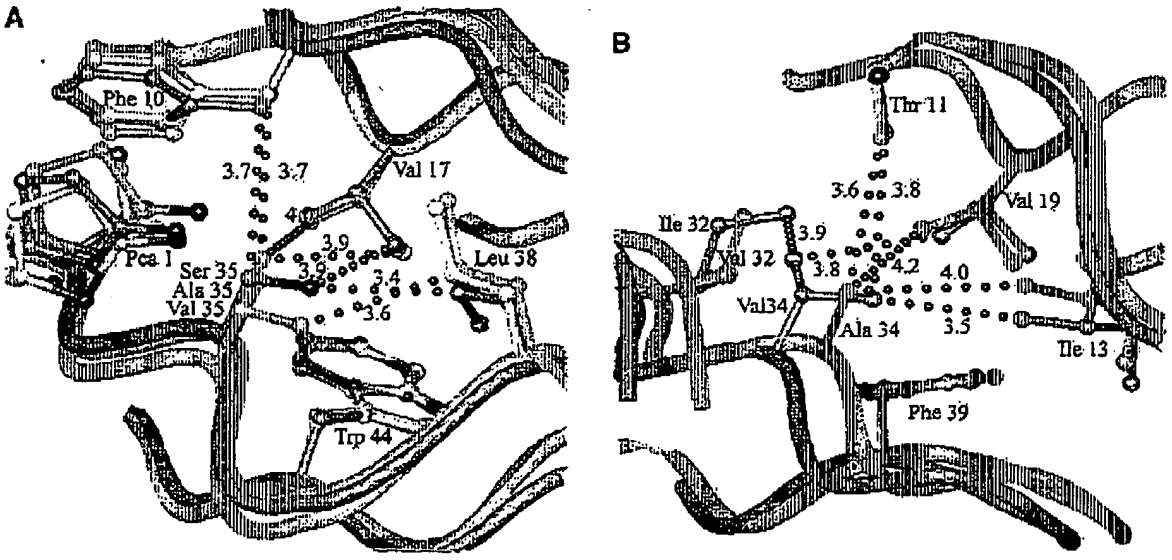


**Figure 4.** (*A*) Interactions and conformational changes close to residue 35 of the fungal GH 12 enzymes from *T. reesei* (wild type and A35V have carbon atoms colored yellow and goldenrod), and *H. schweinitzii* (carbons colored gold). Red bubbles indicate contacts in *T. reesei* A35V Cel12A, blue bubbles in *H. schweinitzii* Cel12A. (*B*) Interactions and conformational changes close to residue 34 of the bacterial GH 12 enzymes from *S. sp. 11AG8* (carbons colored gold), and *S. lividans* (carbons colored yellow). Red bubbles indicate contacts in *S. lividans* CelB2, blue bubbles in *S. sp. 11AG8* Cel12A.

range from 3.4 to 3.9 Å) (Fig. 4A). There are two other substitutions in this region of the structure, W7Y in the loop between β-strands A1/B1 and S63V on β-strand B3 that cause a reduction in the $T_m$ by 1°C and 0.8 °C, respectively, when introduced into *T. reesei* Cel12A. The $T_m$ decrease caused by the S63V substitution is probably due to local rearrangement to alleviate a close contact. Modeling the observed valine side-chain conformation of the *H. schweinitzii* enzyme on the *T. reesei* enzyme gives a 3.0 Å clash between the Cγ2 of V63 and the Cβ of A42. In the *H. schweinitzii* structure, we observe a movement resulting in a distance between the Cγ2 of V63 and the Cβ of A42 of 3.9 Å.

Interestingly, there are three substitutions in this region of the less stable *H. schweinitzii* Cel12A enzyme G41A, and T11S/T16I that cause a $T_m$ increase (2.5°C and 1.1°C, respectively) when introduced in *T. reesei* Cel12A (Table 1b). The $T_m$ increase of 2.5°C caused by the G41A substitution is likely due to a new hydrophobic interaction between the Cβ of A41 and the Cγ2 of residue T214 on β-strand A4. There might also be a contribution from the reduction of the chain entropy in the unfolded form of the protein. The increased $T_m$ caused by the T11S/T16I substitutions is harder to explain, as both of these residues sit in the bottom of the substrate-binding cleft on β-strands B1 and B2 with their functional groups pointing out into the solution. Moreover, they do not have any obvious interactions with other residues that could stabilize the structure.

### T. reesei *Cel12A A35V structure*

The *T. reesei* Cel12A A35V mutant enzyme crystallizes with six NCS-related protein molecules in the asymmetric unit, making up three pairs of interacting molecules. This crystal packing is similar to the wild-type structure (Sandgren et al. 2001). The pairs of molecules interact with an approximate twofold symmetry. Three aromatic side chains from each monomer that are important for substrate binding (W7, W22, and Y111) are buried in this quasi-dimer interface. The indole rings of W7 and W22 are close to the quasi-twofold axis, but do not stack. The side chain of Y111 packs against the edge of the β-sandwich, between the reverse turns at residues 6 and 25. The three pairs are not identical, but show differences that are presumably a result of the crystal packing.

Because of the relatively high resolution of the study, we get insights into conformational variation and potential crystal contact artifacts. The A35V mutant crystallizes in the same space groups as the wild-type enzyme, but with some changes in the unit cell constants that arise from slightly different crystal packing. The biggest differences between the six different NCS molecules correspond to loop regions that take on different conformations in the different NCS molecules, and these are affected by the crystal contacts. Of the loops helping to form the substrate-binding site, those at residues 153, 111, 55, and 26 are most affected, and show clustering into two or more preferred conformations. On the convex side of the β-sheet, the biggest variation occurs at residues 141 and 40. Both are dominated by one outlier and, in the case of the loop at residue 40, the density of the outlier is of rather poor quality. The wild-type enzyme also shows loop-flexing motions, but with a smaller variation as a result of the crystal packing.

The introduction of a valine at position 35 influences packing. Residue A35 is located on β-strand A2 on the smaller β-sheet A close to the amino terminus (Fig. 1A,B). The functional group of the residue points into the core between the interacting β-sheets, where it interacts with a number of spatially adjacent side chains. Residue 35 does not cause a major conformational change or local rearrangement in the structure, however. The side chain takes on a preferred rotamer conformation ($\chi = +180°$) and the methyl groups make good van der Waals contacts with the neighboring hydrophobic residues' side chains (F10, V17, L38, with contacts in the range from 3.6 to 3.9 Å) (Fig. 4A). In the wild-type enzyme, the packing of the alanine 35 Cβ methyl group is not so tight (Fig. 4A). The approximately equivalent pairs of Cαs, from the *T. reesei* Cel12A wild-type and A35V structures can be superimposed with pair-wise RMSD's in the range from 0.4 to 0.6 Å.

### S. sp. 11AG8 *Cel12A structure*

The cd of the bacterial *S. sp. 11AG8* Cel12A enzyme crystallizes with only one molecule in the asymmetric unit. The *S. sp. 11AG8* Cel12A cd structure is most similar to the *Streptomyces lividans* GH 12 enzyme CelB2, with which it shares 72% sequence identity (Fig. 5), and there are no insertions or deletions in the sequences. In the Cel12A cd, there is no observable density for the residues in the linker region beyond Ala 222. The complete set of Cα atoms from the two structures can be superimposed with a RMSD of 0.49 Å.

The *S. sp. 11AG8* Cel12A cd structure has good stereochemistry with only five nonglycine residues outside of a stringent boundary Ramachandran definition (Kleywegt and Jones 1996a). There are two disulfide bridges in the structure. The one between Cys 5 and Cys 31, linking β-strands A1 and A2, is conserved throughout the whole GH 12 (Sandgren et al. 2001). The second disulfide bridge is formed between Cys 64 and Cys 69, and is found in an extended loop between β-strands B3 and A5 that is part of the substrate-binding cleft of the enzyme. Structural alignment shows that the bacterial enzymes have an insertion in this loop compared with the fungal GH 12 enzymes (Sandgren et al. 2001). This suggests that the purpose of this disulfide may be to stabilize the insertion.

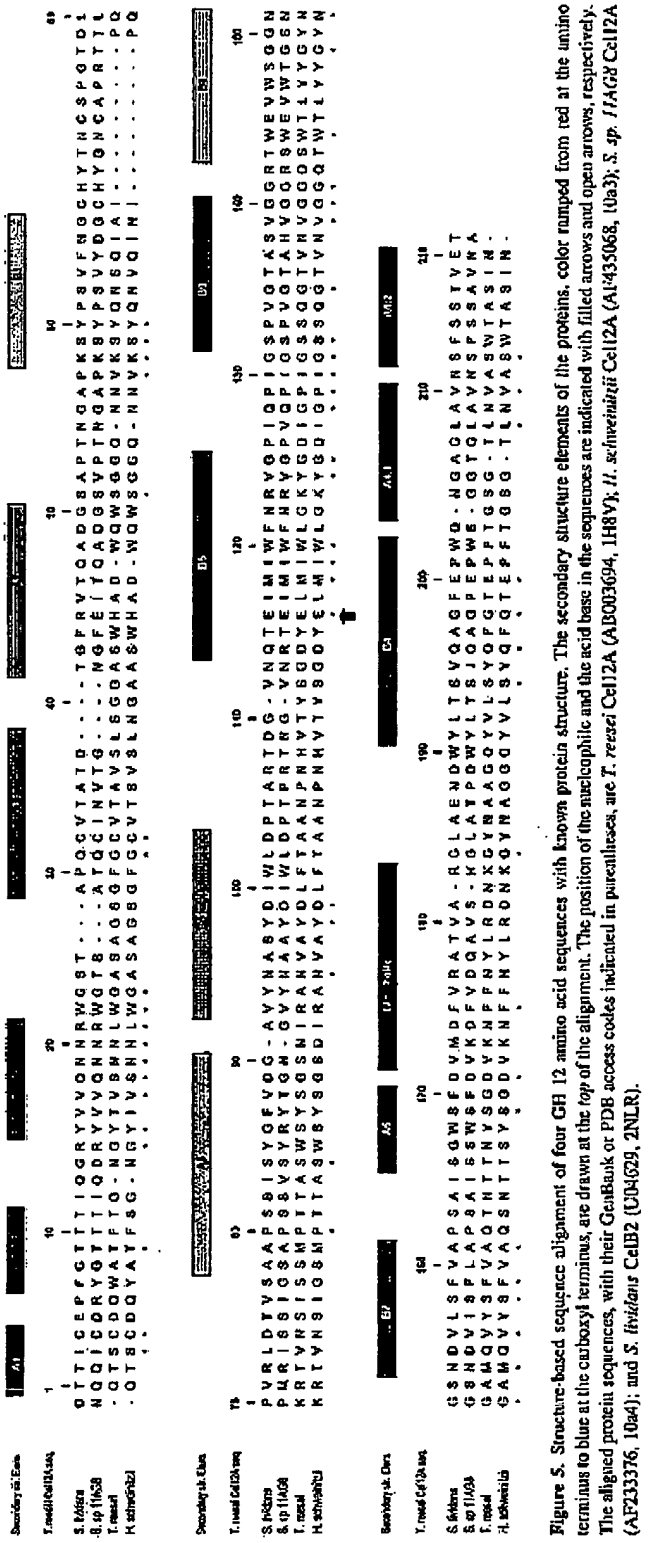

**Figure 5.** Structure-based sequence alignment of four GH 12 amino acid sequences with known protein structure. The secondary structure elements of the proteins, color ramped from red at the amino terminus to blue at the carboxyl terminus, are drawn at the *top* of the alignment. The position of the nucleophile and the acid base in the sequences are indicated with filled arrows and open arrows, respectively. The aligned protein sequences, with their GenBank or PDB access codes indicated in parentheses, are *T. reesei* Cel12A (AB003694, 1H8V); *H. schweinitzii* Cel12A (AI435068, 10a3); *S. sp. 11AG8* Cel12A (AF233376, 10a4); and *S. lividans* CelB2 (U04629, 2NLR).

Two CelB2 structures have been described; in one, the methionine M122 (*S. sp. 11AG8* Cel12A numbering) at the active site has been oxidized (Sulzenbacher et al. 1997), whereas the other structure is in complex with a fluorocellotrioside that has been designed to trap a covalent intermediate to the active site nucleophile (Sulzenbacher et al. 1999). The oligosaccharide has been modeled in two conformations, one as the covalent intermediate to the nucleophile E120, and the other as free substrate. The nucleophile has also been modeled in two conformations in the complex structure. In our bacterial structure, the active site nucleophile is most like the one in the major conformation described for the CelB2 structure with a free fluorocellotrioside bound in the binding cleft. The hydrogen-bond interactions are essentially identical, including the close contact of the acidic groups of the nucleophile and D104 (2.6 Å separation), and the conformations of the ring of identical supporting side chains (Sandgren et al. 2001). The residues lining the substrate-binding groove are highly conserved and very similar in structure. Some of the largest differences involve side-chain shifts of ~0.5 Å for the aromatic rings of W24 in the −2 site, and F39 and Y66 in the −3 site. These differences may be due to substrate binding in CelB2. The *cis*-peptide at P75 is also conserved between the two structures.

One of the largest main-chain differences is at residue A/V34, the equivalent site of the A35V mutation that we show affects temperature stability in the *T. reesei* Cel12A enzyme. The loop corresponding to residues 33–38 has a small rigid body shift, so that in the CelB2 enzyme it moves closer to the upper sheet (Fig. 4B). As with the *T. reesei* Cel12A enzyme, residue 34 is located on β-strand A2 at one edge on the β-sandwich as shown in Figure 1, A and B. In CelB2, the Cβ of A34 has hydrophobic contacts with the three equivalent inward-pointing residues from the upper sheet (T11, I13, and V19, with interatomic separations in the range from 3.6 to 4.2 Å). In the *S. sp. 11AG8* Cel12A structure, the methyl groups of V34 also remain in good van der Waals contact with the equivalent side chains (separations 3.5–4.3 Å), (Fig. 4B). The effect of sequence differences at this residue has only a small effect on the accessibility of solvent to the inner, more closely packed β-sandwich.

## Discussion

Glycoside hydrolase family 12 members show a wide variation in their thermal stability. The $T_m$ variation from the least to the most stable enzyme examined is 20.9°C (Table 1a). These types of stability differences between homologous proteins are common, and there has long been an interest in identifying the sources of the increased stability in the extremophiles (Jaenicke 2000). There have been cases in which, for instance, changes in loop length or compactness, disulfides, electrostatics, and hydrophobic interactions have been shown to contribute to the stability differences among homologs (Jaenicke and Bohm 1998; Vieille et al. 2001). However, no generically reliable rules have emerged to predict stabilizing changes from comparisons of sequences alone or, even modeled structures. Although we find a high level of structural homology between the Cel12A proteins from the mesophile (*T. reesei*) and the more stable enzyme from the alkalophilic bacterium (*S. sp. 11AG8*), it is by no means obvious, given the low sequence identity of 28%, which changes are needed to stabilize the *T. reesei* protein. However, we were able to use the characterization of a less stable homolog, together with examination of the *T. reesei* Cel12A structure and sequence analysis, to guide our choice of recruitment from the more stable homolog.

Cel12A from the fungus *H. schweinitzii* has a $T_m$ (49.2°C) that is 5.2°C lower than the $T_m$ (54.4°C) of the *T. reesei* Cel12A. As the *H. schweinitzii* homolog differed from the *T. reesei* enzyme at only 14 amino acids, it was possible to make all of these changes in individual *T. reesei* Cel12A variants. Mutating the A35 in the *T. reesei* enzyme to a serine, the equivalent amino acid in *H. schweinitzii* Cel12A, results in a drop of the $T_m$ by 4.0°C (Table 1b). An examination of a sequence alignment of the homologous, fungal, and bacterial Cel12A enzymes (Goedegebuur et al. 2002) shows a statistical preference for alanine or valine at that position. The Cel12A from *S. sp. 11AG8* has a valine at that position, and has a $T_m$ (66.8°C) that is 12.4°C higher than that of the *T. reesei* Cel12A enzyme. This analysis, and examination of the *T. reesei* Cel12A structure, led us to recruit the A35V substitution from *S. sp. 11AG8* Cel12A. Mutating this single residue to a valine results in a 7.7°C increase in $T_m$, to a value that is only 4.7°C less than that for the *S. sp. 11AG8* Cel12A enzyme. The effects of the A35S and A35V mutations account for most of the $T_m$ difference between wild-type *T. reesei* Cel12A and the GH 12 homolog from which they were recruited. Thus, a clear finding from this effort was that the buried amino acid at position 35 plays a key role in modulating the stability of these enzymes.

The GH 12 homologs showed a ninefold range in their specific activities on oNPC. In contrast to their effects on stability, the recruitment of the A35S or A35V into *T. reesei* Cel12A has little effect on specific activity. The stability of the GH 12 proteins to thermal denaturation was reflected in the effects of temperature on their activity. As expected, the homolog with greatest thermal stability, Cel12A *S. sp. 11AG8*, showed a continual increase in activity over the temperature range from 25°C V to 60°C (Fig. 3). The less stable *H. schweinitzii* Cel12A enzyme showed a decrease in activity at the highest temperature, presumably due to thermal inactivation. The recruitment of the A35V and A35S into *T. reesei* Cel12A mimicked the behavior of the parent enzymes, with A35V showing a high-temperature activity

benefit compared with wild type, or to the least stable variant A35S.

Examination of the Cel12A structures (Fig. 4A) provides rationalization for the effects on stability. An alanine, serine, or valine at this position has little effect on the backbone structure or on the solvent accessibility of interacting residues. However, when this residue is a serine, it introduces a hydrophilic residue into the hydrophobic core between two β-sheets. When it is a valine, it results in a tighter set of van der Waals interactions with three neighboring side chains in particular, and closing entry to the β-sheet sandwich core. The stabilizing effect of improved interior packing has been shown in some other proteins (Matsumura et al. 1988; Golovanov et al. 2000; Ohmura et al. 2001).

After the completion of this study, the structure of a highly thermostable Cel12 endoglucanase from the thermophilic bacterium *Rhodothermus marinus* was solved (Crennell et al. 2002). On the basis of a purely structural comparison with the two other Cel12 structures then available for *T. reesei* Cel12A (Sandgren et al. 2001) and *S. lividans* CelB2 (Sulzenbacher et al. 1997), the authors suggested that the features that may be largely responsible for the extreme stability of the *R. marinus* Cel12A are an increased number of surface ion pairs, and reduced mobility of a specific loop. In contrast to the CelB2 structure, this 153–158 loop in our bacterial Cel12 structure, *S. sp. 11AG8* Cel12A, has continuous density and is not particularly mobile. The loop's main-chain temperature factors (ca. 21 $Å^2$) are similar to the reported values for the 157–161 loop in *R. marinus* Cel12A (Crennell et al. 2002), less than 1.5 times the average (15.6 $Å^2$) and not greater than those for any other flexible loop in the protein. Similarly, the equivalent 149–154 loop in *T. reesei* Cel12A has temperature factors comparable with those for the loop in the *S. sp. 11AG8* Cel12A structure. None of the mutations introduced in this study would be expected to affect either the number of ion pairs or this specific loop. The *R. marinus* Cel12A protein has only 28% identity with *T. reesei* Cel12A, and fits best in the same Cel12 subfamily (12–3), as all of the *Streptomyces* members (Goedegebuur et al. 2002). It is interesting to note that the amino acid at the position equivalent to A35 in *T. reesei* Cel12A is an alanine for five 12–3 members. However, it is a valine in the remaining two, *R. marinus* Cel12A and *S. sp. 11AG8* Cel12A, both of which show increased thermal stability.

The clear identification of a single residue responsible for large differences in stability within a structural family may be unusual. It is obvious from other studies (Lehmann and Wyss 2001; Lehmann et al. 2002) and from the limited sampling presented in this study that not all recruitments from more stable homologs are stabilizing. For example, the W7Y, S63V, S77G, T163S, and N167S variants of *T. reesei* Cel12A all represent recruitments of the side chains also found in *S. sp. 11AG8* Cel12A, but none were stabilizing (Table 1b). Conversely, not all recruitments from the less-stable homolog were destabilizing. In contrast to A35S, 6 of the 14 substitutions recruited from the *H. schweinitzii* Cel12A stabilized *T. reesei* Cel12A. Recruiting mutants with increased stability from less-stable homologs has been reported previously in other families (Shaw et al. 1999; Lehmann et al. 2000; Perl et al. 2000). There is another indication of the complexity of protein stability even in this relatively simple comparison of two very close (structural and sequence) homologs; summing the $\Delta T_m$'s observed for each of the recruitments from *H. schweinitzii* Cel12A into the *T. reesei* protein gives a net value close to zero, but the *H. schweinitzii* enzyme has a $T_m$ that is 5.2°C lower than the *T. reesei* enzyme (Table 1a). In the well-studied case of T4 lysozyme, the effects of point mutations on stability were shown to be additive (Zhang et al. 1995). The lack of simple additivity seen in this study may be due partly to the spatial clustering of eight of the changes on the first six β-strands, leading to some interdependence of their effects. There may also be some added complexity arising from the fact that the thermal denaturation is irreversible, so that stability is being measured as apparent $T_m$ values rather than as the preferred thermodynamic free energy values. *T. reesei* Cel12A is known to denature reversibly in urea (Arunachalam and Kellis, Jr., 1996) but, despite extensive efforts, we could not find conditions under which we could compare the reversible denaturation of this wild-type protein with that of variants of different stability or of other Cel12 homologs.

In this work, structural and stability studies of closely related proteins provide insight on how specific residues contribute to protein stability in glycoside hydrolase family 12. Several stabilizing mutations have been identified and one expectation for future work is that these could be combined to give greater increases in stability. We are also continuing these studies to include other members of this family.

## Materials and methods

### *Protein expression and purification*

The DNA encoding for *T. reesei* Cel12A was amplified from a cDNA clone (Ward et al. 1993) by use of PCR primers that introduced a *Bgl*II restriction endonuclease site at the 5′ end of the Cel12A gene (immediately upstream of the first ATG codon) and an *Xba*I site at the 3′ end (immediately downstream of the stop codon). The amplified fragment was then subcloned as a *Bgl*II–*Xba*I fragment into pUC19 vector. All mutants were made in this plasmid by use of QuikChange mutagenesis methods (Stratagene) as single-point mutations, with the exception of the double-mutant T11S/T16I, as examination of a preliminary structure suggested that these side chains could interact. The variant or wild-type gene was then subcloned into the *Aspergillus* expression vector pGPT-*pyr*G and the sequence-verified vector was transformed into *A. niger* (Berka and Barnett 1989). The resultant strain was grown in shake flasks or a fermentor. The culture supernatants were treated

overnight with 0.18 mg/mL of endoglucanase H at 37°C. After deglycosylation, ammonium sulfate was added to a final concentration of 0.5 M, and the supernatants were centrifuged. The *T. reesei* Cel12A proteins were then purified from these supernatants by column chromatography. Approximately 1 mL of Butyl Sepharose 4 Fast Flow resin (Amersham Biosciences) per 10 mg of Cel12A protein was loaded and equilibrated in a disposable drip column with 0.05 M Bis Tris Propane, 0.05 M ammonium acetate (pH 8), containing 0.5 M ammonium sulfate. The supernatants were loaded and the column washed with 3 vol of equilibration buffer, before elution with 3 vol of the same buffer without ammonium sulfate. Each column volume was collected as a separate fraction with the pure Cel12A proteins appearing in the second elution fraction. The identity of the purified proteins was confirmed by tryptic mapping by HPLC Mass Spectrometry (data not shown).

The genes encoding the *T. koningii* Cel12A, *H. schweinitzii* Cel12A, *G. roseum* Cel12C, and *F. javanicum* Cel12A endoglucanases were cloned, expressed in *A. niger* (Goedegebuur et al. 2002), and purified essentially as described for the *T. reesei* Cel12A, with the following changes in the chromatography: The EndoH-treated cell-free supernatants of the first three were loaded and washed in the presence of 1 M ammonium sulfate; *T. koningii* Cel12A enzyme was purified by hydrophobic interaction chromatography (HIC) using the Butyl Sepharose 4 Fast Flow resin; *H. schweinitzii* Cel12A was purified on an Octyl Sepharose 4 Fast Flow column, and *G. roseum* Cel12C was purified on a Phenyl Sepharose 6 Fast Flow column; *F. javanicum* Cel12A was purified by ion-exchange chromatography on Poros 20 HQ (Applied Biosystems) in 0.02 M $NaH_2PO_4$/NaOH (pH 6.8), using a 20-column volume gradient from 0 to 1 M NaCl.

The *S. sp. 11AG8* Cel12A enzyme was expressed heterologously in *Streptomyces lividans* (US patent no. 6287839). Ultrafiltered and concentrated fermentation broth was first treated with 10% $Na_2SO_4$ (w/v) to remove some contaminants by a slow, overnight precipitation at 37°C, followed by centrifugation. *S. sp. 11AG8* Cel12A was precipitated from the supernatant by adding solid $(NH_4)_2SO_4$ to 1.3 M, and centrifuging after holding the solution at 4°C for 1 h. The pellets were solubilized in ~1/10 of the original volume of 0.1 M sodium acetate, 0.02% Tween 20 (pH 5.5). The solubilized pellet was diluted 20-fold with 0.02 M $NaH_2PO_4$/NaOH, 1.0 M $(NH_4)_2SO_4$ at pH 6.8 (HIC equilibration buffer), and loaded onto an Octyl Sepharose 4 Fast Flow column. The column was washed with a 3-column volume of equilibration buffer, and then eluted with a step of buffer containing 0.5 M $(NH_4)_2SO_4$ (pH 6.8). Early elution fractions containing both the Cel12A cd, and some high molecular weight contaminants (not characterized) were not pooled. The most pure fractions (≥97% pure) were pooled and then concentrated and diafiltered into 0.1 M ammonium acetate (pH 5.5). The final concentration to ~15 mg/mL was done using Millipore Ultrafree Centrifugal Filter Devices with Biomax 10,000 Dalton molecular weight cut-off membranes. The apparent molecular weight of the purified protein on an SDS polyacrylamide gel was 25–26 kD. This is the molecular weight expected for the *S. sp. 11AG8* Cel12A cd (US patent no. 6287839). The carboxy-terminal cellulose-binding module (CBM) is removed by proteolysis of the linker between the cd and the CBM, resulting in two forms of the cd, 230 and 233 residues in length.

### *Thermal denaturation experiments*

CD experiments were performed on an Aviv 62ADS spectrophotometer (Protein Solutions), equipped with a five-position thermoelectric cell holder supplied by Aviv. Buffer conditions were 0.05 M Bis-Tris propane, 0.05 M ammonium acetate adjusted to pH 8.0 with acetic acid. The final protein concentration for each experiment was in the range from 10 to 20 μM. Data was collected in a 0.1-cm path length cell. The thermal denaturation experiments were performed at 217 nm, the wavelength in the far-UV spectra with maximum signal difference, as expected for a predominantly β-sheet protein. The temperature was increased from 30 to 90°C with data collected every 2°. The equilibration time at each temperature was 0.1 min and data were collected for 4 sec per sample. The thermal denaturation data were fitted to a two-state transition (Chen et al. 1992) using Savuka software provided by Dr. Osman Bilsel (University of Massachusetts Medical School, N. Worcester, MA). The mid-point of the transition ($T_m$) is an apparent value, because the thermal denaturation of all the Cel12A proteins studied was not reversible. Because of this and a slow kinetic component due to aggregation, sufficiently fast scan rates were chosen, and experiments were carefully controlled for time spent in the thermal denaturation transition, so that small temporal variations did not lead to apparent $T_m$ differences.

The pH of maximum thermal stability for all Cel12A homologs studied was pH 5 (data not shown) as reported for the urea denaturation of *T. reesei* Cel12A (Arunachalam and Kellis, Jr., 1996). A suboptimal condition of pH 8.0 was used to facilitate detection of thermal stability differences.

### *Specific enzyme activity*

To evaluate specific enzyme activity, and to monitor expression and purification of all the Cel12 proteins, an oNPC (Sigma N 4764) hydrolysis assay was used. In a microtiter plate, 100 μL 50 mM sodium acetate (pH 5.5) and 20 μL 25 mg/mL oNPC in assay buffer was added. Once equilibrated, 10 μL of cellulase was added and the plate incubated at 40°C for 10 min, unless otherwise specified. To stop the reaction, 70 μL of 0.2 M glycine (pH 10.0) was added. The plate was then read in a microtiter plate reader at 410 nm. As a reference, 10 μL of a 0.1 mg/mL solution of *T. reesei* Cel12A enzyme provided an OD of around 0.3. The concentration of the *T. reesei* Cel12A enzyme was determined by the absorbance at 280 nm, using an extinction coefficient of 78711 $M^{-1}$ $cm^{-1}$ (3.352 $g/L^{-1}$), determined experimentally by the method of Edelhoch as described in Pace et al. (1995). Extinction coefficients for the other Cel12 homologs were calculated on the basis of their amino acid compositions (Pace et al. 1995).

### *Protein crystallization*

The *T. reesei* Cel12A A35V variant was crystallized under conditions similar to the wild-type enzyme (Sandgren et al. 2001) using 200 mM cacodylate buffer (pH 6.0), 200 mM calcium acetate and 10%–30% (w/w) mono-methyl-ether (mme) PEG 2000, at 20–24°C with hanging and sitting drops (McPherson 1982). Crystallization drops were prepared by mixing equal amounts of protein solution (20 mg/mL) and crystallization agent. Large single, wedge-shaped crystals grew to a maximum size of 1 mm in all directions within 1–2 d. The crystals belong to the space group $P2_1$ with cell dimensions a = 68.3 Å, b = 71.3 Å, c = 119.3 Å, and β = 91.5°, and have a calculated $V_m$ of 1.9 (Matthews 1968) with six molecules in the asymmetric unit.

The ammonium acetate concentration in the *S. sp. 11AG8* Cel12A protein sample was reduced to 0.01 M by dilution with water, and the enzyme reconcentrated to 15 mg/mL by ultrafiltration. Crystallization experiments were set up using the vapor-diffusion method (McPherson 1982) with sitting drops at 25°C. A total of 5 μL of precipitant was added to 5 μL of protein solution

in a micro-bridge (Hampton Research), suspended over 0.5 mL of reagent in the well of a VDX crystallization plate, and sealed with a cover-slide. Crystals were obtained with the Hampton research crystal screen I (Jancarik and Kim 1991). After optimizing the conditions, large single crystals (1 × 1 × 0.5 mm) were obtained from 10%–20% (w/w) mme PEG 5000, 200 mM sodium cacodylate (pH 5.0–6.0). The crystals belong to space group $P2_12_12_1$ with cell dimensions a = 65.1 Å, b = 54.5 Å, c = 62.5 Å, and have a calculated $V_m$ of 2.4 (Matthews 1968) with one molecule in the asymmetric unit.

*H. schweinitzii* Cel12A crystals were obtained from a crystallization agent containing 200 mM cacodylate buffer (pH 6.0), 200 mM ammonium acetate, 10%–30% (w/w) mme PEG 2000, using the vapor-diffusion method with hanging and sitting drops (McPherson 1982), at 20–24°C. Crystallization drops were prepared by mixing equal amounts of protein solution (15 mg/mL) and crystallization agent to a final drop-size of 6–8 µL. Initially, bunches of elongated crystals grew to a maximum size of 0.1 × 0.01 × 0.01 mm within 1 to 2 d. To slow down the crystallization process, 1% isopropyl alcohol was added to the crystallization solution. Single, larger crystals (0.1 mm in all directions) were obtained after 6 mo of incubation. These crystals belong to space group $P2_1$ with the following cell dimensions: a = 62.5 Å, b = 77.5 Å, c = 83.4 Å, and β = 98.5°, and have a calculated $V_m$ of 2.0 (Matthews 1968), with four molecules in the asymmetric unit.

### *X-ray data collection*

All crystals were equilibrated in 25%–40% mme PEG 2000, 200 mM sodium cacodylate (pH 5.0), mounted in a cryo-loop, plunge frozen in liquid nitrogen (prior to transportation to the synchrotron), or flash frozen in a nitrogen cryo-stream mounted on the detector. All data sets were collected from single crystals and at 100 K. Data collection and processing statistics for the structures are given in Table 2a.

The *S. sp. 11AG8* Cel12A data set was processed with the software distributed with the R-axis-II image plate detector (Molecular Structure Corp). The *H. schweinitzii* and *T. reesei* A35VCel12A data sets were processed and scaled with DENZO and SCALEPACK (Otwinowski and Minor 1997).

All subsequent data processing, after image integration, were performed using the CCP4 package (Collaborative Computational Project Number 4 1994) unless otherwise stated.

### *Structure solution and refinement*

The bacterial *S. sp. 11AG8* Cel12A cd structure was solved by MR, using X-plor 3.1 (Brünger 1992b), with the bacterial *S. lividans* CelB catalytic core structure (PDB code 1NLR) as the search model. The resulting map clearly revealed the differences between the search model and the new structure. The initial structure model was built in Xfit, part of the Xtalview suite of programs (McRee 1999) and refined in X-plor 3.1 (Brünger 1992b). The last two rounds of model building and refinement were performed with alternating cycles of O (Jones et al. 1991) and Refmac 5.0 (Murshudov et al. 1997).

The structure of *H. schweinitzii* Cel12A was solved by MR with Amore (Navaza 1994), using all of the homologous wild-type *T. reesei* Cel12A structure (residues 1–218) as a search model. The MR search gave one clear solution with four molecules in the asymmetric unit. The initial map, calculated using the phases from the MR solution, was improved by exploiting the fourfold NCS of the asymmetric unit. Initial local symmetry operators were determined in O (Jones et al. 1991), refined with RAVE (Kleywegt and Jones 1999), and the electron densities were cyclically averaged using the program DM (Cowtan and Main 1998). This produced an easily interpretable electron density map, allowing the unambiguous building of a complete molecule, including all residues from 1 to 218. A single NAG residue could be clearly identified covalently bound to Asn 164. Strict NCS constraints between the four molecules in the asymmetric unit were only used in the first rounds of refinement. No restraints were applied in the final rounds.

The changes in unit cell parameters of the *T. reesei* Cel12A A35V mutant compared with the wild-type structure required the use of molecular replacement, using Amore (Navaza 1994) to solve the structure. Structure model building and refinement of the A35V mutant and the *H. schweinitzii* Cel12A structures were performed with alternating cycles of model building using O (Jones et al. 1991), and maximum likelihood model refinement, bulk solvent corrections, anisotropic scaling, and automated bulk solvent correction using CNS version 1.0 (Brünger et al. 1998). In the final rounds of structure refinement, Refmac 5.0 was used (Murshudov et al. 1997).

All water molecules in the structure models were located automatically by using the water-picking protocols in the refinement programs. These water molecules were then selected manually or discarded by inspection. A set of reflections from the data sets were used to monitor the R-free (Brünger 1992a). PROCHECK (Laskowski et al. 1993) was used to evaluate the stereo-chemical

**Table 2b.** *Structure refinement and final model statistics*

| GH 12 Protein | *T. reesei* A35V | *S. sp. 11AG8* wild type | *H. schweinitzii* wild type |
|---|---|---|---|
| PDB access code | 1oa2 | 1oa3 | 1oa4 |
| Resolution used in refinement (Å) | 20–1.5 | 29–1.5 | 29–1.70 |
| Reflections in: | | | |
| working set | 167,265 | 25,464 | 83,042 |
| test set | 5174 | 821 | 2573 |
| R & $R_{free}$ factor (%) | 20.1; 21.9 | 18.1; 19.3 | 19.2; 21.7 |
| Protein molecules in AU | 6 | 1 | 4 |
| Protein atoms | 9978 | 1686 | 6648 |
| Water molecules | 643 | 417 | 6648 |
| ⟨B⟩ (Å²) | 20.4 | 16.3 | 11.2 |
| Protein ⟨B⟩ (Å²) | 20.2 | 15.6 | 10.5 |
| water ⟨B⟩ (Å²) | 24.5 | 25.8 | 19.0 |
| RMSD bond lengths from ideal (Å)[a] | 0.018 | 0.0119 | 0.011 |
| RMSD bond angles from ideal (°)[a] | 1.7 | 1.5 | 1.3 |
| RMSD ΔB on bonded atoms (Å²) | 0.9 | 1.0 | 1.2 |
| RMSD NCS Cα (Å) | 0.5 | — | 0.2 |
| RMSD NCS all atoms (Å) | 0.7 | — | 0.5 |
| Stringent Ramachandran outliers (%)[b] | 0.8 | 2.0 | 0.9 |

Values were calculated with O (Jones et al. 1991; Jones and Kjeldgaard 1997), CNS 1.0 (Brünger et al. 1998), MOLEMAN (Kleywegt and Jones 1996b), LSQMAN (Kleywegt and Jones 1997), Refmac 5.0 (Murshudov et al. 1997).

[a] From Engh and Huber (1991).

[b] According to the stringent boundary definition of Kleywegt and Jones (1996a).

quality of the final models. Summaries of model refinement statistics are given in Table 2b.

All structural comparisons were made with O (Jones et al. 1991), and figures were prepared with O and rendered with MolRay (Harris and Jones 2001). Coordinates and structure-factor amplitudes have been deposited with the Protein Data Bank (Bernstein et al. 1977), and have access codes ID's 1oa2, 1oa3, and 1oa4 for the *T. reesei* A35V, *H. schweinitzii*, and *S. sp. 11AG8* Cel12A enzymes, respectively. The final electron density maps are available for viewing as part of the electron density server (EDS) service at http://ashtray.bmc.uu.se/eds, with ID's 1oaA2, 1oa3, and 1oa4.

## Acknowledgments

We thank Dr. Osman Bilsel and Dr. Dave Lambright for providing an early version of the Savuka software; Christian Paech and Sigrid Paech for the tryptic mapping of Cel12A proteins; and Dr. Roopa Ghimikar for critical reading of the manuscript. This work was supported in part by a subcontract from the Department of Energy's Office of Fuels Development.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## References

Arunachalam, U. and Kellis, Jr., J.T. 1996. Folding and stability of endoglucanase III, a single-domain cellulase from *Trichoderma reesei*. *Biochemistry* 35: 11379–11385.

Berka, R.M. and Barnett, C.C. 1989. The development of gene expression systems for filamentous fungi. *Biotechnol. Adv.* 7: 127–154.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, Jr., E.T., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112: 535–542.

Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Moncur, V., Nitz, M., Rose, D.R., Tull, D., Wakarchuck, W.W., et al. 1998. Mechanisms of cellulases and xylanases. *Biochem. Soc. Trans.* 26: 156–160.

Brünger, A.T. 1992a. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472–475.

———. 1992b. *X-PLOR version 3.1: A system for X-ray crystallography and NMR*. Yale University Press, New Haven, CT.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography and NMR system (CNS): A new software suite for macromolecular structure determination. *Acta Crystallogr.* D54: 905–921.

Chen, B.L., Baase, W.A., Nicholson, H., and Schellman, J.A. 1992. Folding kinetics of T4 lysozyme and nine mutants at 12 degrees C. *Biochemistry* 31: 1464–1476.

Collaborative Computational Project Number 4. 1994. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr.* D50: 760–763.

Cowtan, K. and Main, P. 1998. Miscellaneous algorithms for density modification. *Acta Crystallogr.* D54: 487–493.

Crennell, S.J., Hreggvidsson, G.O., and Nordberg Karlsson, E. 2002. The structure of *Rhodothermus marinus* Cel12A, a highly thermostable family 12 endoglucanase, at 1.8 Å resolution. *J. Mol. Biol.* 320: 883–897.

Engh, R.A. and Huber, R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.* A47: 392–400.

Goedegebuur, F., Fowler, T., Phillips, J., Van Der Kley, P., Van Solingen, P., Dankmeyer, L., and Power, S.D. 2002. Cloning and relational analysis of 15 novel fungal endoglucanases from family 12 glycosyl hydrolase. *Curr. Genet.* 41: 89–98.

Golovanov, A.P., Vergoten, G., and Arseniev, A.S. 2000. Stabilization of proteins by enhancement of inter-residue hydrophobic contacts: Lessons of T4 lysozyme and barnase. *J. Biomol. Struct. Dyn.* 18: 477–491.

Harris, M. and Jones, T.A. 2001. Molray—A web interface between O and the POV-Ray ray tracer. *Acta Crystallogr.* D57: 1201–1203.

Henrissat, B. and Davies, G.J. 2000. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol.* 124: 1515–1519.

Jaenicke, R. 2000. Stability and stabilization of globular proteins in solution. *J. Biotechnol.* 79: 193–203.

Jaenicke, R. and Bohm, G. 1998. The stability of proteins in extreme environments. *Curr. Opin. Struct. Biol.* 8: 738–748.

Juncarik, J. and Kim, S.H. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J. Appl. Crystallogr.* 24: 409–411.

Jones, T.A. and Kjeldgaard, M.O. 1997. Electron-density map interpretation. *Methods Enzymol.* 277: 173–208.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* A47: 110–119.

Khademi, S., Zhang, D., Swanson, S.M., Wartenberg, A., Witte, K., and Meyer, E.F. 2002. Determination of the structure of an endoglucanase from *Aspergillus niger* and its mode of inhibition by palladium chloride. *Acta Crystallogr.* D58: 660–667.

Kleywegt, G.J. and Jones, T.A. 1996a. Phi/Psi-chology: Ramachandran revisited. *Structure* 4: 1395–1400.

———. 1996b. xdlMAPMAN and xdlDATAMAN: Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr.* D52: 826–828.

———. 1997. Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* 277: 525–545.

———. 1999. Software for handling macro molecular envelopes. *Acta Crystallogr.* D55: 941–944.

Laskowski, R.A., MacArthur, M.V., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26: 283–291.

Lehmann, M. and Wyss, M. 2001. Engineering proteins for thermostability: The use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol.* 12: 371–375.

Lehmann, M., Pasamontes, L., Lassen, S.F., and Wyss, M. 2000. The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta* 1543: 408–415.

Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S.F., Pasamontes, L., van Loon, A.P., and Wyss, M. 2002. The consensus concept for thermostability engineering of proteins: Further proof of concept. *Protein Eng.* 15: 403–411.

Matsumura, M., Becktel, W.J., and Matthews, B.W. 1988. Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* 334: 406–410.

Matthews, B.W. 1968. Solvent content of protein crystals. *J. Mol. Biol.* 33: 491–497.

McPherson, A.J. 1982. *Preparation and analysis of protein crystals*. John Wiley and Sons, New York.

McRee, D.E. 1999. XtalView/Xfit: A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* 125: 156–165.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* D53: 240–255.

Navaza, J. 1994. AMoRe: An automated package for molecular replacement. *Acta Crystallogr.* A50: 157–163.

Ohmura, T., Ueda, T., Ootsuka, K., Saito, M., and Imoto, T. 2001. Stabilization of hen egg white lysozyme by a cavity-filling mutation. *Protein Sci.* 10: 313–320.

Okada, H., Mori, K., Tada, K., Nogawa, M., and Morikawa, Y. 2000. Identification of active site carboxylic residues in *Trichoderma reesei* endoglucanase Cel12A by site-directed mutagenesis. *J. Mol. Catalysis B*: 249–255.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276: 307–326.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* 4: 2411–2423.

Perl, D., Mueller, U., Heinemann, U., and Schmid, F.X. 2000. Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol.* 7: 380–383.

Sandgren, M., Shaw, A., Ropp, T.H., Wu, S., Bott, R., Cameron, A.D., Ståhlberg, J., Mitchinson, C., and Jones, T.A. 2001. The X-ray crystal structure of the *Trichoderma reesei* family 12 endoglucanase 3, Cel12A, at 1.9 Å resolution. *J. Mol. Biol.* 308: 295–310.

Schulein, M. 1997. Enzymatic properties of cellulases from *Humicola insolens*. *J. Biotechnol.* **57:** 71–81.

Shaw, A., Bott, R., and Day, A.G. 1999. Protein engineering of α-amylase for low pH performance. *Curr. Opin. Biotechnol.* **10:** 349–352.

Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C., and Davies, G.J. 1997. The *Streptomyces lividans* family 12 endoglucanase: Construction of the catalytic core, expression, and X-ray structure at 1.75 Å resolution. *Biochemistry* **36:** 16032–16039.

Sulzenbacher, G., Mackenzie, L.F., Wilson, K.S., Withers, S.G., Dupont, C., and Davies, G.J. 1999. The crystal structure of a 2-fluorocellotriosyl complex of the *Streptomyces lividans* endoglucanase CelB2 at 1.2 Å resolution. *Biochemistry* **38:** 4826–4833.

van Solingen, P., Meijer, D., van der Kleij, W.A., Barnett, C., Bolle, R., Power, S.D., and Jones, B.E. 2001. Cloning and expression of an endocellulase gene from a novel *Streptomycete* isolated from an East African soda lake. *Extremophiles* **5:** 333–341.

Vieille, C., Sriprapundh, D., Kelly, R.M., and Zeikus, J.G. 2001. Xylose isomerases from Thermotoga. *Methods Enzymol.* **330:** 215–224.

Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Clarkson, K., and Bott, R. 1993. *The Tricell 93 symposium*, Espoo, Finland.

Zhang, X.J., Baase, W.A., Shoichet, B.K., Wilson, K.P., and Matthews, B.W. 1995. Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. *Protein Eng.* **8:** 1017–1022.

Zechel, D.L., He, S., Dupont, C., and Withers, S.G. 1998. Identification of Glu-120 as the catalytic nucleophile in *Streptomyces lividans* endoglucanase celB. *Biochem. J.* **336:** 139–145.